Appendix B. Source and Accuracy of Estimates

SOURCE OF DATA

Most estimates in this report come from data obtained in October 1989 in the Current Population Survey (CPS). The Bureau of the Census conducts the survey every month, although this report uses only October data for its estimates. The October survey uses two sets of questions, the basic CPS and the supplement.

Basic CPS. The basic CPS collects primarily labor force data about the civilian noninstitutional population. Interviewers ask questions concerning labor force participation about each member 15 years old and over in every sample household.

The present CPS sample was selected from the 1980 decennial census files with coverage in all 50 States and the District of Columbia. The sample is continually updated to account for new residential construction. It's located in 729 areas and includes 1,973 counties, independent cities, and minor civil divisions. About 56,100 occupied housing units are eligible for interview every month. Interviewers are unable to obtain interviews at about 2,500 of these units because the occupants are not found at home after repeated calls or are unavailable for some other reason.

Since the introduction of the CPS, the Bureau of the Census has redesigned the CPS sample several times to improve the quality and reliability of the data and to satisfy changing data needs. The most recent changes were completely implemented in July 1985.

The following table summarizes changes in the CPS designs for the years for which data appear in this report.

Description of the Current Population Survey

Time period	Number of sample areas		
		Interviewed	Not interviewed
1989	729	53,600	2,500
1984	629	59,000	2,500

October Supplement. In addition to the basic CPS questions, interviewers asked supplementary questions in October about the use and ownership of computers.

Estimation Procedure. This survey's estimation procedure inflates weighted sample results to independent estimates of the civilian noninstitutional population of the United States by age, sex, race and Hispanic/non-Hispanic categories. The independent estimates were based on statistics from the 1980 Census of Population; statistics on births, deaths, immigration and emigration; and statistics on the size of the Armed Forces. The independent population estimates include some, but not all, undocumented immigrants.

ACCURACY OF ESTIMATES

Since the CPS estimates come from a sample, they may differ from figures from a complete census using the same questionnaires, instructions, and enumerators. A sample survey estimate has two possible types of error: sampling and nonsampling. The accuracy of an estimate depends on both types of error, but the full extent of the nonsampling error is unknown. Consequently, one should be particularly careful when interpreting results based on a relatively small number of cases or on small differences between estimates. The standard errors for CPS estimates primarily indicate the magnitude of sampling error. They also partially measure the effect of some nonsampling errors in responses and enumeration, but do not measure systematic biases in the data. (Bias is the average over all possible samples of the differences between the sample estimates and the desired value.)

Nonsampling Variability. Nonsampling errors can be attributed to many sources. These sources include the inability to obtain information about all cases in the sample, definitional difficulties, differences in the interpretation of questions, respondents' inability or unwillingness to provide correct information or to recall information, errors made in data collection such as in recording or coding the data, errors made in processing the data, errors made in estimating values for missing data, and failure to represent all units with the sample (undercoverage).

CPS undercoverage results from missed housing units and missed persons within sample households. Compared to the level of the 1980 Census, overall CPS undercoverage is about 7 percent. CPS undercoverage varies with age, sex, and race. Generally, undercoverage is larger for males than for females and larger for Blacks and other races combined than for Whites. As described previously, ratio estimation to independent

Table B-3. **Standard Errors of Estimated Percentages: Black and Other Races**

Base of estimated percentage (thousands)	Estimated percentage							
	1 or or 99	2 or or 98	5 or 95	10 or 90	20 or 80	25 or 75	35 or 65	50
10	6.1	8.5	13.3	18.7	24.4	26.4	29.1	30.5
25	3.8	5.4	8.4	11.6	15.4	16.7	18.4	19.3
50	2.7	3.8	5.9	8.2	10.9	11.8	13.0	13.6
100	1.9	2.7	4.2	5.8	7.7	8.3	9.2	9.6
250	1.2	1.7	2.7	3.7	4.9	5.3	5.8	6.1
500	0.9	1.2	1.9	2.6	3.4	3.7	4.1	4.3
1,000	0.6	0.8	1.3	1.8	2.4	2.6	2.9	3.0
5,000	0.3	0.4	0.6	0.8	1.1	1.2	1.3	1.4
10,000	0.2	0.3	0.4	0.6	0.8	0.8	0.9	1.0
15,000	0.2	0.2	0.3	0.5	0.6	0.7	0.8	0.8
25,000	0.12	0.2	0.3	0.4	0.5	0.5	0.6	0.6

Note: For a particular characteristic, see table B-5 for the appropriate factor to apply to the above standard errors.
For regional estimates, multiply the above standard errors by 0.91, 0.97, 0.99, and 1.17 for the Northeast, Midwest, South, and West, respectively.

Here x is the total number of persons, families, households, or unrelated individuals in the base of the percentage, p is the percentage ($0 \leqslant p \leqslant 100$), and b is the parameter in table B-5 associated with the characteristic in the numerator of the percentage.

Illustration. Table 1 shows that of the 24,007,000 households with children from 6 to 17 years of age, 5,998,000, or 25.7 percent, have a computer in the household. Using formula (3), the appropriate factor from table B-5 (1.0), and a standard error from table B-2, the approximate standard error is 1.0x0.4 = 0.4.

From table B-5, the appropriate b parameter is 4,651. Using formula (4), the approximate standard error of 25.7 percent is

$$s_{x,p} = \sqrt{\frac{4{,}651}{24{,}007{,}000}\,25.7(100 - 25.7)} = 0.6$$

This means that the 90-percent confidence interval for the percentage of households with children from 6 to 17 with a computer in the household is from 24.7 to 26.7 percent (i.e., 25.7 ± 1.6x0.6).

Standard Error of a Difference. The standard error of the difference between two sample estimates is approximately equal to

$$s_{x\ y} = \sqrt{s_x^2 + s_y^2} \qquad (5)$$

where s_x and s_y are the standard errors of the estimates, x and y. The estimates can be numbers, percentages, ratios, etc. This will represent the actual standard error quite accurately for the difference between estimates of the same characteristic in two different areas, or for the difference between separate and uncorrelated characteristics in the same area. However, if there is a high positive (negative) correlation between the two characteristics, the formula will overestimate (underestimate) the true standard error.

Illustration. Table 1 shows there were 1,430,000 households with a computer and with incomes from $15,000

Table B-4. **Standard Errors of Estimated Percentages: Hispanic**

Base of estimated percentage (thousands)	Estimated percentage							
	1 or 99	2 or 98	5 or 95	10 or 90	20 or 80	25 or 75	35 or 65	50
10	8.0	11.3	17.6	24.3	32.4	35.0	38.6	40.5
25	5.1	7.2	11.2	15.4	20.5	22.2	24.4	25.6
50	3.6	5.1	7.9	10.9	14.5	15.7	17.3	18.1
100	2.6	3.6	5.6	7.7	10.2	11.1	12.2	12.8
250	1.6	2.3	3.5	4.9	6.5	7.0	7.7	8.1
500	1.1	1.6	2.5	3.4	4.6	5.0	5.5	5.7
1,000	0.8	1.1	1.8	2.4	3.2	3.5	3.9	4.0
5,000	0.4	0.5	0.8	1.1	1.4	1.6	1.7	1.8
10,000	0.2	0.4	0.6	0.8	1.0	1.1	1.2	1.3
15,000	0.2	0.3	0.5	0.6	0.8	0.9	1.0	1.0

Note: For a particular characteristic, see table B-5 for the appropriate factor to apply to the above standard errors.
For regional estimates, multiply the above standard errors by 0.91, 0.97, 0.99, and 1.17 for the Northeast, Midwest, South, and West, respectively.

Table B-5. **Standard Error Parameters and Factors: October 1989 and October 1984**

Characteristic	Parameters, October 1989			Parameters, October 1984		
	a	b	f	a	b	f
PERSONS						
Total, White, and non-Hispanic:						
Enrolled in school	-0.000012	2,743	1.0	-0.000010	2,311	0.9
Household type, age of householder, presence of children	-0.000023	4,651	1.3	-0.000019	3,918	1.2
Unemployed	-0.000020	2,619	1.0	-0.000017	2,206	0.9
Black and other:						
Enrolled in school	-0.000107	3,711	1.0	-0.000090	3,125	0.9
Household type, age of householder, presence of children	-0.000279	6,672	1.3	-0.000235	5,620	1.2
Unemployed	-0.000185	3,010	0.9	-0.000156	2,536	0.8
Hispanic:						
Enrolled in school	-0.000320	6,551	1.0	-0.000149	2,558	0.6
Household type, age of householder, presence of children	-0.000550	11,244	1.3	-0.000029	4,961	0.9
Unemployed	-0.000260	3,154	0.7	-0.000108	2,087	0.6
FAMILIES AND HOUSEHOLDS						
Total, White, and non-Hispanic:						
Household type, age of householder, presence of children	-0.000013	1,846	0.8	-0.000011	1,555	0.8
Household income	-0.000013	2,287	0.9	-0.000011	1,927	0.8
Employment status and occupation of householder	-0.000033	2,390	0.9	-0.000031	2,013	0.9
Unemployed	-0.000020	2,619	1.0	-0.000017	2,206	0.9
Black and other:						
Household type, age of householder, presence of children	-0.000116	1,668	0.7	-0.000097	1,405	0.6
Household income	-0.000118	2,493	0.8	-0.000100	2,100	0.8
Employment status and occupation of householder	-0.000294	2,390	0.8	-0.000247	2,013	0.7
Unemployed	-0.000185	3,010	0.9	-0.000156	2,536	0.8
Hispanic:						
Household type, age of householder, presence of children	-0.000195	2,811	0.7	-0.000022	1,592	0.5
Household income	-0.000199	4,201	0.8	-0.000016	2,709	0.6
Employment status and occupation of householder	-0.000207	2,503	0.6	-0.000121	2,086	0.6
Unemployed	-0.000260	3,154	0.7	-0.000108	2,087	0.6

Note: For regional estimates, multiply the a and b parameters by 0.83, 0.93, 0.99, and 1.37 for the Northeast, Midwest, South, and West, respectively. The 1984 parameters have been revised since the 1984 computer usage report was published.

to $19,999. It also shows there were 1,066,000 households with a computer and with incomes from $20,000 to $24,999. The apparent difference is 364,000. Using formula (2) and the appropriate parameters from table B-5, the approximate standard errors of these estimates are

$$\sqrt{-0.000013 \times 1{,}430{,}000^2 + 2{,}287 \times 1{,}430{,}000} = 57{,}000$$

and

$$\sqrt{-0.000013 \times 1{,}066{,}000^2 + 2{,}287 \times 1{,}066{,}000} = 49{,}000$$

respectively.

Therefore, from formula (5), the approximate standard error of the estimated difference of 364,000 households is

$$\sqrt{57{,}000^2 + 49{,}000^2} = 75{,}000.$$

This means that the 90-percent confidence interval for the difference between households with a computer and incomes from $15,000 to $19,999 and households with a computer and incomes from $20,000 to $24,999 is from 244,000 to 484,000 (i.e., 364,000 $\pm$ 1.6x75,000).

Therefore, a conclusion that the average estimate of the difference, derived from all possible samples, lies within a range computed in this way would be correct for roughly 90 percent of all possible samples. Since this interval does not contain zero, we can conclude with 90 percent confidence that the number of households with a computer and incomes from $15,000 to $19,999 is greater than the number of households with computers and incomes from $20,000 to $24,999.